

IN THE CLAIMS:

The following is a complete listing of the claims now pending. This listing replaces all earlier versions and listings of the claims.

Claim 1 (previously presented): An apparatus for identifying one or more portions of data in a database for comparison with a query input by a user, the query and the portions of data each comprising a sequence of sub-word units, said apparatus comprising:

a memory for storing data defining a plurality of sub-word unit classes, each class comprising sub-word units that are confusable with other sub-word units in the same class;

a memory for storing an index having a plurality of entries, each entry having an associated identifier for identifying the entry and each entry comprising:

a key associated with the entry and which is related to the identifier for the entry in a predetermined manner; and

a number of pointers which point to portions of data in the database which correspond to the key associated with the entry,

wherein each key comprises a sequence of sub-word unit classifications which is derived from a corresponding sequence of sub-word units appearing in the database by classifying each of the sub-word units in the sequence into one of the plurality of sub-word unit classes;

means for classifying each of the sub-word units in the input query into one of the plurality of sub-word unit classes and for defining one or more sub-sequences of query sub-word unit classifications;

means for determining a corresponding identifier for an entry in the index for each of the one or more sub-sequences of query sub-word unit classifications;

means for comparing the key associated with each of the determined identifiers determined by said determining means with the corresponding sub-sequence of query sub-word unit classifications; and

means for retrieving one or more pointers from the index in accordance with the output of said comparing means, which one or more pointers identify the one or more portions of data in the database for comparison with the input query.

Claim 2 (previously presented): An apparatus according to claim 1, wherein the sub-word units are phonemes.

Claim 3 (original): An apparatus according to claim 1, wherein at least ten sub-word unit classes are defined in advance.

Claim 4 (original): An apparatus according to claim 1, wherein each key is related to the corresponding identifier by a predetermined mathematical function.

Claim 5 (original): An apparatus according to claim 4, wherein each key is related to the corresponding identifier by the following equation:

$$\left(\prod_{i=1}^W [C[i]K_c] \right) \text{Mod} S,$$

where K_c is the number of sub-word unit classes, S is the number of entries in the index, C is the number of the sub-word class to which the i^{th} sub-word unit in the sequence of sub-word units corresponding to the key belongs, and W is the number of sub-word unit classifications in each key.

Claim 6 (previously presented): An apparatus according to claim 1, wherein said determining means is operable to identify a new identifier for another entry in the index for a subsequence of query sub-word unit classifications if said comparing means determines that the key for the identifier is not the same as the subsequence of query sub-word unit classifications.

Claim 7 (previously presented): An apparatus according to claim 6, wherein said determining means is operable to determine a new identifier using the following equation:

$$IDX = \text{Mod } S,$$

where IDX is the identifier, S is the number of entries in the index, and V is a predetermined number.

Claim 8 (previously presented): An apparatus according to claim 1, wherein the key for one or more of the entries is a null key indicating that there are no pointers stored in the index for that entry.

4b Claim 9 (previously presented): An apparatus according to claim 4, wherein said determining means is operable to determine a corresponding identifier for each subsequence of query sub-word unit classifications using the predetermined mathematical function.

Claim 10 (previously presented): An apparatus according to claim 1, wherein the input query is a typed query and said apparatus further comprises means for converting the typed query into the sequence of sub-word units.

Claim 11 (previously presented): An apparatus according to claim 1, wherein the input query is a spoken query and said apparatus further comprises a speech recognition system for processing the spoken query and for outputting the sequence of sub-word units.

Claim 12 (previously presented): An apparatus for searching a database in response to a query input by a user, the database comprising a plurality of sequences of sub-word units and the query comprising at least one sequence of sub-word units, said apparatus comprising:

an apparatus according to any of claims 1 to 11 for identifying one or more portions of data in the database for comparison with the input query; and

means for comparing the one or more sequences of query sub-word units with the identified one or more portions of data in the database.

Claim 13 (previously presented): An apparatus according to claim 12, wherein said means for comparing the input query with the portions of data in the database uses a dynamic programming comparison technique.

Claim 14 (previously presented): An apparatus according to claim 12, further comprising means for retrieving one or more data files in accordance with the results of said comparing means.

Claim 15 (previously presented): An apparatus for identifying one or more portions of data in a database for comparison with a query input by a user, the query and the portions of data each comprising a sequence of features, said apparatus comprising:

a memory for storing data defining a plurality of feature classes, each class comprising features that are confusable with other features in the same class;

a memory for storing an index having a plurality of entries, each entry having an associated identifier for identifying the entry and each entry comprising:

a key associated with the entry and which is related to the identifier for the entry in a predetermined manner; and

a number of pointers which point to portions of data in the database which correspond to the key associated with the entry,

wherein each key comprises a sequence of feature classifications which is derived from a corresponding sequence of features appearing in the database by classifying each of the features in the sequence into one of the plurality of feature classes;

means for classifying each of the features in the input query into one of the plurality of feature classes and for defining one or more sub-sequences of query feature classifications;

means for determining a corresponding identifier for an entry in the index for each of the one or more sub-sequences of query feature classifications;

means for comparing the key associated with each of the determined identifiers determined by said determining means with the corresponding sub-sequence of query feature classifications; and

means for retrieving one or more pointers from the index in accordance with the output of said comparing means, which one or more pointers identify the one or more portions of data in the database for comparison with the input query.

Claim 16 (canceled)

Claim 17 (previously presented): A method of identifying one or more portions of data in a database for comparison with a query input by a user, the query and the portions of data each comprising a sequence of sub-word units, the method comprising the steps of:

storing data defining a plurality of sub-word unit classes, each class comprising sub-word units that are confusable with other sub-word units in the same class;

storing an index having a plurality of entries, each entry having an associated identifier for identifying the entry,

a key associated with the entry and which is related to the identifier for the entry in a predetermined manner, and

4b a number of pointers which point to portions of data in the database which correspond to the key associated with the entry,

wherein each key comprises a sequence of sub-word unit classifications which is derived from a corresponding sequence of sub-word units appearing in the database by classifying each of the sub-word units in the sequence into one of the plurality of sub-word unit classes;

classifying each of the sub-word units in the input query into one of the plurality of sub-word unit classes and for defining one or more sub-sequences of query sub-word unit classifications;

determining a corresponding identifier for an entry in the index for each of the one or more sub-sequences of query sub-word unit classifications;

comparing the key associated with each of the determined identifiers determined in said determining step with the corresponding sub-sequence of query sub-word unit classifications; and

retrieving one or more pointers from the index in accordance with the output of said comparing step, which one or more pointers identify the one or more portions of data in the database for comparison with the input query.

Claim 18 (previously presented): A method according to claim 17, wherein the sub-word units are phonemes.

Claim 19 (original): A method according to claim 17, wherein at least ten sub-word unit classes are defined in advance.


Claim 20 (original): A method according to claim 17, wherein each key is related to the corresponding identifier by a predetermined mathematical function.

Claim 21 (previously presented): A method according to claim 20, wherein each key is related to the corresponding identifier by the following equation:

$$\left(\prod_{i=1}^w [C[i]K_c] \right) \text{Mod} S$$

where K_c is the number of sub-word unit classes, S is the number of entries in the index, C is the number of the sub-word class to which the i^{th} sub-word unit in the sequence of sub-word units corresponding to the key belongs, and W is the number of sub-word unit classifications in each key.

Claim 22 (previously presented): A method according to claim 17, wherein said determining step identifies a new identifier for another entry in the index for a subsequence of query sub-word unit classifications if said comparing step determines that the key for the identifier is not the same as the subsequence of query sub-word unit classifications.

 Claim 23 (previously presented): A method according to claim 22, wherein said determining step determines a new identifier using the following equation:

$$IDX = Mod S$$

where IDX is the identifier, S is the number of entries in the index, and V is a predetermined number.

Claim 24 (previously presented): A method according to claim 17, wherein the key for one or more of the entries is a null key indicating that there are no pointers stored in the index for that entry.

Claim 25 (previously presented): A method according to claim 23, wherein said determining step determines a corresponding identifier for each subsequence of query sub-word unit classifications using the predetermined mathematical function.

Claim 26 (previously presented): A method according to claim 17, wherein the input query is a typed query and wherein the method further comprises a step of converting the typed query into the sequence of sub-word units.

B Claim 27 (previously presented): A method according to claim 17, wherein the input query is a spoken query and wherein the method further comprises a step of using a speech recognition system to process the spoken query to generate the sequence of sub-word units.

Claim 28 (previously presented): A method of searching a database in response to a query input by a user, the database comprising a plurality of sequences of sub-word units and the query comprising at least one sequence of sub-word units, said method comprising:

method steps of claim 17 for identifying one or more portions of data in the database for comparison with the input query; and

a step of comparing the one or more sequences of query sub-word units with the identified one or more portions of data in the database.

Claim 29 (previously presented): A method according to claim 28, wherein said comparing step uses a dynamic programming comparison technique to compare the input query with the portions of data.

Claim 30 (previously presented): A method according to claim 28, further comprising a step of retrieving one or more data files in dependence upon the results of said comparing step.

Claims 31 and 32 (canceled)

Claim 33 (previously presented): An apparatus for identifying one or more portions of data in a database for comparison with a query input by a user, the query and the portions of data each comprising a sequence of sub-word units, said apparatus comprising:

an index having a plurality of entries, each of which includes a key comprising a sequence of sub-word unit classifications, which key is derived from a corresponding sequence of sub-word units appearing in the database by classifying each of the sub-word units in the sequence into one of a plurality of sub-word unit classes, each class comprising sub-word units that are confusable with other sub-word units in the same class; and

identifying means operable to identify the one or more portions of data in the database for comparison with the input query, using keys in the index which correspond to sub-word unit classifications in the input query.

Claim 34 (previously presented): An apparatus according to claim 4, wherein each key is, related to the corresponding identifier by the following equation:

$$\left(\sum_{i=1}^n K_c^{(n-1)} \cdot C[i] \right) \text{Mod } S$$

where K_c is the number of sub-word unit classes, S is the number of entries in the index, C is the number of the sub-word class to which the i^{th} sub-word unit in the sequence of sub-word units corresponding to the key belongs, and n is the number of sub-word unit classifications in each key.

Claim 35 (previously presented): A method according to claim 20, wherein each key is related to the corresponding identifier by the following equation:

$$\left(\sum_{i=1}^n K_c^{(n-i)} \cdot C[i] \right) \text{Mod } S$$

where K_c is the number of sub-word unit classes, S is the number of entries in the index, C is the number of the sub word class to which the i^{th} sub-word unit in the sequence of sub-word units corresponding to the key belongs, and n is the number of sub-word unit classifications in each key.

Claim 36 (previously presented): An apparatus for identifying one or more portions of data in a database for comparison with a query input by a user, the query and the portions of data each comprising a sequence of sub-word units, the apparatus comprising:

a first memory operable to store data defining a plurality of sub-word unit classes, each class comprising sub-word units that are confusable with other sub-word units in the same class;

a second memory operable to store an index having a plurality of entries, each entry having an associated identifier for identifying the entry and each entry comprising:

a key associated with the entry and which is related to the identifier for the entry in a predetermined manner; and

a number of pointers which point to portions of data in the database which correspond to the key for the entry,

wherein each key comprises a sequence of sub-word unit classifications which is derived from a corresponding sequence of sub-word units appearing in the database by classifying each of the sub-word units in the sequence into one of the plurality of sub-word unit classes;

a classifier operable to classify each of the sub-word units in the input query into one of the plurality of sub-word unit classes and to define one or more sub-sequences of query sub-word unit classifications;

a determiner operable to determine a corresponding identifier for an entry in the index for each of the one or more sub-sequences of query sub-word unit classifications;

a comparator operable to compare the key associated with each of the determined identifiers determined by said determiner with the corresponding sub-sequence of query sub-word unit classifications; and

a retriever operable to retrieve one or more pointers from the index in accordance with the output of said comparator, which one or more pointers identify the one or more portions of data in the database for comparison with the input query.

*b*¹ Claim 37 (previously presented): An apparatus according to claim 36 wherein the sub-word units are phonemes.

Claim 38 (previously presented): An apparatus according to claim 36 wherein at least ten sub-word unit classes are defined in advance.

Claim 39 (previously presented): An apparatus according to claim 36 wherein each key is related to the corresponding identifier by a predetermined mathematical function.

Claim 40 (previously presented): An apparatus according to claim 39, wherein each key is related to the corresponding identifier by the following equation:

$$\left(\prod_{i=1}^W [C[i]K_c] \right) \text{Mod} S$$

where K_c is the number of sub-word unit classes, S is the number of entries in the index, C is the number of the sub-word class to which the i^{th} sub-word unit in the sequence of sub-word units corresponding to the key belongs, and W is the number of sub-word unit classifications in each key.

Claim 41 (previously presented): An apparatus according to claim 39, wherein each key is related to the corresponding identifier by the following equation:

$$\left(\sum_{i=1}^n K_c^{(n-1)} \cdot C[i] \right) \text{Mod } S$$

P
where K_c is the number of sub-word unit classes, S is the number of entries in the index, C is the number of the sub-word class to which the i^{th} sub-word unit in the sequence of sub-word units corresponding to the key belongs, and n is the number of sub-word unit classifications in each key.

Claim 42 (previously presented): An apparatus according to claim 36, wherein said determiner is operable to identify a new identifier for another entry in the index for a subsequence of query sub-word unit classifications if said comparator determines that the key for the identifier is not the same as the subsequence of query sub-word unit classifications.

Claim 43 (previously presented): An apparatus according to claim 42, wherein said determiner is operable to determine a new identifier using the following equation:

$$IDX = \text{Mod } S$$

where IDX is the identifier, S is the number of entries in the index, and V is a predetermined number.

Claim 44 (previously presented): An apparatus according to claim 36, wherein the key for one or more of the entries is a null key indicating that there are no pointers stored in the index for that entry.

b' Claim 45 (previously presented): An apparatus according to claim 39, wherein said determiner is operable to determine a corresponding identifier for each subsequence of query sub-word unit classifications using the predetermined mathematical function.

Claim 46 (previously presented): An apparatus according to claim 36, wherein the input query is a typed query and said apparatus further comprises a converter operable to convert the typed query into the sequence of sub-word units.

Claim 47 (previously presented): An apparatus according to claim 36, wherein the input query is a spoken query and said apparatus further comprises a speech recognition system operable to process the spoken query and to output the sequence of sub-word units.

Claim 48 (previously presented): An apparatus for searching a database in response to a query input by a user, the database comprising a plurality of sequences of sub-word units and the query comprising at least one sequence of sub-word units, said apparatus comprising:

an apparatus according to claim 36 operable to identify one or more portions of data in the database for comparison with the input query; and

a sequence comparator operable to compare the one or more sequences of query sub-word units with the identified one or more portions of data in the database.

Claim 49 (previously presented): An apparatus according to claim 48, wherein said sequence comparator is operable to compare the input query with the portions of data in the database using a dynamic programming comparison technique.

Claim 50 (previously presented): An apparatus according to claim 48, further comprising a data file retriever operable to retrieve one or more data files in accordance with the results of said sequence comparator.

Claim 51 (previously presented): An apparatus for identifying one or more portions of data in a database for comparison with a query input by a user, the query and the portions of data each comprising a sequence of features, said apparatus comprising:

a first memory operable to store data defining a plurality of feature classes, each class comprising features that are confusable with other features in the same class;

a second memory operable to store an index having a plurality of entries, each entry having an associated identifier for identifying the entry and each entry comprising:

a key associated with the entry and which is related to the identifier for the entry in a predetermined manner; and

a number of pointers which point to portions of data in the database which correspond to the key for the entry,

wherein each key comprises a sequence of feature classifications which is derived from a corresponding sequence of features appearing in the database by classifying each of the features in the sequence into one of the plurality of feature classes;

40 a classifier operable to classify each of the features in the input query into one of the plurality of feature classes and to define one or more sub-sequences of query feature classifications;

a determiner operable to determine a corresponding identifier for an entry in said index for each of said one or more sub-sequences of query feature classifications;

a comparator operable to compare the key associated with each of the determined identifiers determined by said determiner with the corresponding sub-sequence of query feature classifications; and

a retriever operable to retrieve one or more pointers from the index in accordance with the output of said comparator, which one or more pointers identify the one or more portions of data in the database for comparison with the input query.

Claim 52 (previously presented): An index for use in searching a database, the index having a plurality of entries, each entry having an associated identifier for identifying the entry and at least one entry comprising:

a key which is related to the corresponding identifier in a predetermined manner; and

one or more pointers which point to locations within the database corresponding to the key for the entry;

wherein each key comprises a sequence of sub-word unit classifications which is derived from a corresponding sequence of sub-word units appearing in the database by classifying each of the sub-word units in the sequence into one of a plurality of sub-word unit classes, the sub-word unit classes being defined in advance and each comprising sub-word units that are confusable with other sub-word units in the same class.

Claim 53 (previously presented): A storage medium storing program code for executing a method of controlling a processor to identify one or more portions of data in a database for comparison with a query input by a user, the query and the portions of data each comprising a sequence of sub-word units, said program code comprising:

code for storing data defining a plurality of sub-word unit classes, each class comprising sub-word units that are confusable with other sub-word units in the same class;

code for storing an index having a plurality of entries, each entry having an associated identifier for identifying the entry and each entry comprising:

a key associated with the entry and which is related to the identifier for the entry in a predetermined manner; and

a number of pointers which point to portions of data in the database which correspond to the key for the entry,

wherein each key comprises a sequence of sub-word unit classifications which is derived from a corresponding sequence of sub-word units appearing in the database by classifying each of the sub-word units in the sequence into one of the plurality of sub-word unit classes;

code for classifying each of the sub-word units in the input query into one of the plurality of sub-word unit classes and defining one or more sub-sequences of query sub-word unit classifications;

code for determining a corresponding identifier for an entry in the index for each of the one or more sub-sequences of query sub-word unit classifications;

code for comparing the key associated with each of the determined identifiers determined by said determining code with the corresponding sub-sequence of query sub-word unit classifications; and

code for retrieving one or more pointers from the index in accordance with the output by said comparing code, which one or more pointers identify the one or more portion of data in the database for comparison with the input query.

Claim 54 (currently amended): A program for controlling a processor to implement a method of identifying one or more portions of data in a database for comparison with a query input by a user, the query and the portions of data each comprising a sequence of sub-word units, said program comprising:

code for storing data defining a plurality of sub-word unit classes, each class comprising sub-word units that are confusable with other sub-word units in the same class;

code for storing an index having a plurality of entries, each entry having an associated identifier for identifying the entry and each entry comprising:

a key associated with the entry and which is related to the identifier for the entry in a predetermined manner; and

a number of pointers which point to portions of data in the database which correspond to the key for the entry,

wherein each key comprises a sequence of sub-word unit classifications which is derived from a corresponding sequence of sub-word units appearing in the database by classifying each of the sub-word units in the sequence into one of the plurality of sub-word unit classes;

code for classifying each of the sub-word units in the input query into one of the plurality of sub-word unit classes and defining one or more sub-sequences of query sub-word unit classifications;

code for determining a corresponding identifier for an entry in the index for each of the one or more sub-sequences of query sub-word unit classifications;

code for comparing the key associated with each of the determined identifiers determined by said determining code with the corresponding sub-sequence of query sub-word unit classifications; and

code for retrieving one or more pointers from the index in accordance with the output by said comparing code, which one or more pointers identify the one or more portions of data in the database for comparison with the input query.